The ALPPS risk score: Avoiding futile use of ALPPS

Linecker, Michael ; Stavrou, Gregor A ; Oldhafer, Karl J ; Jenner, Robert M ; Seifert, Burkhardt ; Lurje, Georg ; Bednarsch, Jan ; Neumann, Ulf ; Capobianco, Ivan ; Nadalin, Silvio ; Robles-Campos, Ricardo ; de Santibañes, Eduardo ; Malagó, Massimo ; Lesurtel, Mickael ; Clavien, Pierre-Alain ; Petrowsky, Henrik

Abstract: OBJECTIVES To create a prediction model identifying futile outcome in ALPPS (Associating Liver Partition and Portal vein ligation for Staged hepatectomy) before stage 1 and stage 2 surgery. BACKGROUND ALPPS is a 2-stage hepatectomy, which incorporates parenchymal transection at stage 1 enabling resection of extensive liver tumors. One of the major criticisms of ALPPS is the associated high mortality rate up to 20%. METHODS Using the International ALPPS Registry, a risk analysis for futile outcome (defined as 90-day or in-hospital mortality) was performed. Futility was modeled using multivariate regression analysis and a futility risk score formula was computed on the basis of the relative size of logistic model regression coefficients. RESULTS Among 528 ALPPS patients from 38 centers, a futile outcome was observed in 47 patients (9%). The pre-stage 1 model included age 67 years or older [odds ratio (OR) = 5.7], and tumor entity (OR = 3.8 for biliary tumors) as independent predictors of futility from multivariate analysis. For the pre-stage 1 model scores of 0, 1, 2, 3, 4 and 5 were associated with futile risk of 2.7%, 4.9%, 8.6%, 15%, 24%, and 37%. The pre-stage 2 model included major complications (grade 3b) after stage 1 (OR = 3.4), serum bilirubin (OR = 4.4), serum creatinine (OR = 5.4), and cumulative pre-stage 1 risk score (OR = 1.9). The model predicted futility risk of 5%, 10%, 20%, and 50% for patients with scores of 3.9, 4.7, 5.5, and 6.9, respectively. CONCLUSIONS Both models have an excellent prediction to assess the individual risk of futile outcome after ALPPS surgery and can be used to avoid futile use of ALPPS.

DOI: https://doi.org/10.1097/SLA.0000000000001914

The ALPPS Risk Score

Avoiding Futile Use of ALPPS

Michael Linecker, MD, Gregor A. Stavrou, MD,†‡ Karl J. Oldhafer, MD,†‡ Robert M. Jenner, MD,† Burkhardt Seifert, PhD,§ Georg Lurje, MD,¶ Jan Bednarsch, MD,¶ Ulf Neumann, MD,¶ Ivan Capobianco, MD,|| Silvio Nadalin, MD,|| Ricardo Robles-Campos, MD,** Eduardo de Santibañes, MD, PhD, FACS,†† Massimo Malagó, MD,‡‡ Mickael Lesurtel, MD, PhD,* Pierre-Alain Clavien, MD, PhD, FACS,* and Henrik Petrowsky, MD, FACS**

Objectives: To create a prediction model identifying futile outcome in ALPPS (Associating Liver Partition and Portal vein ligation for Staged hepatectomy) before stage 1 and stage 2 surgery.

Background: ALPPS is a 2-stage hepatectomy, which incorporates parenchymal transection at stage 1 enabling resection of extensive liver tumors. One of the major criticisms of ALPPS is the associated high mortality rate up to 20%.

Methods: Using the International ALPPS Registry, a risk analysis for futile outcome (defined as 90-day or in-hospital mortality) was performed. Futility was modeled using multivariate regression analysis and a futility risk score formula was computed on the basis of the relative size of logistic model regression coefficients.

Results: Among 528 ALPPS patients from 38 centers, a futile outcome was observed in 47 patients (9%). The pre-stage 1 model included age 67 years or older [odds ratio (OR) = 5.7], and tumor entity (OR = 3.8 for biliary tumors) as independent predictors of futility from multivariate analysis. For the pre-stage 1 model scores of 0, 1, 2, 3, 4 and 5 were associated with futile risk of 2.7%, 4.9%, 8.6%, 15%, 24%, and 37%. The pre-stage 2 model included major complications (grade $\geq$3b) after stage 1 (OR = 3.4), serum bilirubin (OR = 4.4), serum creatinine (OR = 5.4), and cumulative pre-stage 1 risk score (OR = 1.9). The model predicted futility risk of 5%, 10%, 20%, and 50% for patients with scores of 3.9, 4.7, 5.5, and 6.9, respectively.

Conclusions: Both models have an excellent prediction to assess the individual risk of futile outcome after ALPPS surgery and can be used to avoid futile use of ALPPS.

Keywords: ALPPS, associating liver partition and portal vein ligation for staged hepatectomy, futility, mortality, outcome, risk score

(*Ann Surg* 2016;xx:xxx–xxx)

ALPPS (Associating Liver Partition and Portal vein Ligation for Staged Hepatectomy)[1,2] is a novel variant of 2-stage hepatectomy combining parenchymal transection with portal vein occlusion at stage 1 resulting in accelerated liver hypertrophy and enabling completion hepatectomy (stage 2) within a short period of 1 to 2 weeks. Despite these great advantages, ALPPS has been seriously challenged by criticisms on reported high morbidity and mortality. Even experienced centers have reported high mortality rates up to 20%.[3–7]

Although several risk factors for mortality have been identified,[7–10] there are currently no tools available allowing the quantification of the individual cumulative risk for futile outcome after ALPPS. Therefore, knowing the individual risk before stage 1 or stage 2 surgery would help hepatobiliary surgeons to guide better patient selection and proper surgical treatment to avoid futile use of ALPPS.

The present study was undertaken to analyze the short-term outcome of patients undergoing ALPPS to identify pre-stage 1 and pre-stage 2 risk factors for surgery-related futile outcome, which was defined as 90-day or in-hospital mortality. The main objective was to develop 2 prediction models that allow the quantified assessment of the cumulative risk for futile outcome upfront before stage 1 and subsequently before stage 2 surgery.

PATIENTS AND METHODS

Study Design

All patients entered in the International ALPPS Registry from October 2012 to October 2015 were analyzed. The ALPPS Registry is an international platform with voluntary data entry to collect worldwide the experience of this procedure with currently counting more than 200 participating centers. The Registry is coordinated by the Department of Surgery, University of Zurich, Switzerland using the Web-based data capture system secuTrial (Interactive System, Berlin, Germany) which is accessible through www.alpps.net. Data monitoring was performed by a specialized study nurse and the coordinator of the ALPPS registry on a weekly basis. The primary goal of the study was to develop 2 models to predict surgery-related futile outcome before stage 1 and before stage 2 surgery. Surgery-related futility was defined as 3-month or in-hospital mortality after

From the *Department of Surgery and Transplantation, Swiss HPB and Transplant Center, University Hospital Zurich, Zurich, Switzerland; †Department of General and Abdominal Surgery, Asklepios Hospital Barmbek, Hamburg, Germany; ‡Semmelweis University, Budapest, Campus Hamburg, Germany; §Institute for Epidemiology, Biostatistics and Prevention, University of Zurich, Zurich, Switzerland; ¶Department of General, Visceral and Transplantation Surgery, University Hospital Aachen, RWTH Aachen, Germany; ||General, Visceral and Transplant Surgery, University Hospital Tübingen, Tübingen, Germany; **Department of Surgery and Liver and Pancreas transplantation, Virgen de la Arrixaca Clinic and University Hospital, Murcia, Spain; ††Department of Surgery, Division of HPB Surgery, Liver Transplant Unit, Italian Hospital Buenos Aires, Buenos Aires, Argentina; and ‡‡Department of HPB and Liver Transplant Surgery, University College London, Royal Free Hospital, London, UK.

Reprints: Henrik Petrowsky, MD, FEBS, FACS, Department of Surgery and Transplantation, Swiss HPB and Transplant Center Zurich, University Hospital Zurich, Rämistrasse 100, CH-8091 Zurich, Switzerland. E-mail: Henrik.Petrowsky@usz.ch.

Disclosure: The study was supported by the Clinical Research Priority Program of the University of Zurich as part of the project ''Non-resectable liver tumors''. The authors report no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

ISSN: 0003-4932/14/26105-0821
DOI: 10.1097/SLA.0000000000001914

stage 1 surgery. Ethics approval was obtained from the Cantonal Ethics Committee Zurich and the Registry was registered at ClinicalTrials.gov (NCT01924741). The Scientific Committee of the ALPPS registry approved the present study on March 16, 2015. The data sets of the included patients were exported from the Registry on September 29, 2015. In addition to these data, high-volume centers were asked to provide data outside the Registry. Upon agreement, the data were subsequently incorporated in the ALPPS Registry.

Inclusion and Exclusion Criteria

All eligible Registry patients were screened for inclusion in the study. Only ALPPS procedures with complete or partial parenchymal transection were included; other transection variants such as tourniquet- and ablation-assisted ALPPS[11,12] were not included. Furthermore, only centers which reported 5 or more ALPPS cases were included to minimize the learning effect. Lacking data of the primary end point and failure to proceed with stage 2 surgery were further exclusion criteria.

Definitions of Variables and Outcome Measures

The primary end point of the study was surgery-related futile outcome, which was defined as 90-day or in-hospital mortality calculated from the date of stage 1 surgery. Liver tumors were classified into 3 groups: colorectal liver metastases (CRLM), biliary tumors including cholangiocarcinoma and gallbladder cancer, and noncolorectal/nonbiliary tumors. The Charlson Comorbidity Index was recorded as global measure of comorbidities.[13] Cardiovascular disease (CVD) was defined as one of the following parameters captured in the registry: congestive heart failure, history of myocardial infarction, medical treatment including antianginal agents, aspirin, digoxin, antihypertensives, or other cardiovascular drugs. Complications after ALPPS were recorded using the Clavien-Dindo classification.[14] Major complications were defined as grade 3b or higher.[14] To assess liver failure, 2 commonly used criteria were applied before and after stage 1 surgery: Model of End-stage Liver Disease (MELD)[15] and the International Study Group of Liver Surgery (ISGLS) criteria.[16] Further outcome variables included interstage liver function performance, future liver remnant (FLR) growth, and length of intensive care unit and hospital stay.

Statistical Methods

Descriptive Statistics

Continuous data were expressed as median and interquartile range, categorical data as absolute numbers and percentage. *P* values for comparing outcomes between futile and nonfutile groups were computed using the χ^2 test for categorical variables and the Mann-Whitney *U* test for continuous variables. *P* values less than 0.05 were considered statistically significant.

Multivariate Regression Analysis

All variables, which proofed to be statistically significant in the univariate analysis and had a data completion rate of at least 70%, were considered for multivariate regression analysis (Tables 1 and 2). Regression analysis was separately performed for pre-stage 1 and for pre-stage 2 data to identify risk factors in a pre-stage 1 and pre-stage 2 model. Based on discussion between clinician and biostatistician,

TABLE 1. Pre-stage 1 Characteristics

Variable	Data Completion, %	Futile (n = 47)	Nonfutile (n = 481)	*P*
Age, yr	95	69 (64–74)	61 (52–68)	<0.001
Sex; male, n, %	100	30 (64)	273 (57)	0.397
BMI, kg/m^2	94	26 (23–29)	25 (23–28)	0.481
Liver tumor	95			
CRLM, n, %		19 (41)	324 (71)	<0.001
Biliary tumors, n, %		18 (39)	58 (13)	<0.001
Non-CRLM/nonbiliary, n, %		9 (20)	69 (15)	0.401
Chemotherapy, n, %	93	20 (44)	313 (70)	<0.001
Comorbidities§				
Charlson Comorbidity Index	92	6 (2–6)	6 (6–6)	0.068
Age-adjusted Charlson Comorbidity Index	89	7 (5–9)	7 (6–8)	0.878
CVD, n, %	100	10 (21)	41 (9)	0.009
Renal disease, n, %	100	3 (6)	7 (1)	0.051
Diabetes mellitus, n, %	100	8 (17)	44 (9)	0.117
COPD, n, %	100	3 (6)	16 (3)	0.234
Abdominal surgery, n, %*	100	12 (26)	229 (47)	0.003
Liver baseline characteristics				
FLR/BW pre-stage 1	82	0.33 (0.29–0.41)	0.35 (0.27–0.46)	0.375
sFLR pre-stage 1	85	0.19 (0.16–0.24)	0.21 (0.16–0.28)	0.182
Macrosteatosis, n, %	42	10 (44)	64 (32)	0.351
Fibrosis, n, %	44	6 (26)	52 (25)	1.000
Steatohepatitis, n, %	46	6 (27)	39 (18)	0.261
CASH, n, %	41	5 (26)	40 (20)	0.554
Bilirubin, mg/dL	84	0.70 (0.41–1.58)	0.59 (0.40–0.82)	0.011
INR	79	1.00 (0.99–1.10)	1.00 (0.95–1.10)	0.331
Creatinine, mg/dL	74	0.83 (0.70–1.00)	0.80 (0.70–0.92)	0.336
MELD score	66	7 (6–10)	6 (6–7)	0.054

Categorical variables presented as count and percentage (%), continous variables presented as median and interquartile range (IQR).

*History of previous major abdominal surgery.

§Comorbidities not mentioned were assumed not exisiting.

BMI indicates body mass index; CASH, chemotherapy-associated steatohepatitis; COPD, chronic obstructive pulmonary disease; BW, bodyweight; INR, international normalized ratio; sFLR, standardized FLR.

TABLE 2. Pre-stage 2 Characteristics

Variable	Data Completion, %	Futile (n = 47)	Nonfutile (n = 481)	*P*
Characteristics of stage 1				
Partial ALPPS, n, %	100	2 (4)	43 (9)	0.411
Use of Pringle, n, %	99	16 (35)	125 (29)	0.225
CVP, mmHg	53	5 (3–6)	5 (4–6)	0.331
Transfusion, n, %	95	15 (32)	99 (22)	0.142
Overall complications, n, %	77	27 (60)	102 (28)	<0.001
Major complications, n, %*	85	15 (33)	30 (7)	<0.001
Dialysis after stage 1, n, %	10	9 (100)	3 (7)	<0.001
Operation time stage 1, min	82	341 (270–490)	300 (250–385)	0.026
ICU stay after stage 1, days	69	4 (1–9)	1 (1–3)	<0.001
Hospital stay after stage 1, days	81	12 (9–18)	10 (7–13)	0.003
Interstage liver performance				
Liver volume characteristics				
FLR/BW pre-stage 2	80	0.59 (0.46–0.74)	0.62 (0.51–0.75)	0.139
sFLR pre-stage 2	77	0.33 (0.27–0.44)	0.40 (0.33–0.47)	0.006
Δ sFLR†	73	0.14 (0.08–0.20)	0.16 (0.11–0.22)	0.127
sFLR increase, %	66	77 (52–105)	70 (46–106)	0.615
Liver tests at POD 5				
Serum bilirubin, mg/dL	84	1.13 (0.77–4.05)	0.80 (0.50–1.33)	<0.001
INR	81	1.20 (1.10–1.47)	1.10 (1.00–1.21)	0.008
Serum creatinine, mg/dL	74	1.06 (0.71–1.85)	0.70 (0.59–0.86)	<0.001
MELD score	66	11 (8–20)	8 (7–10)	<0.001
ISGLS, n, %	83	10 (23)	36 (9)	0.015
Liver tests pre-stage 2				
Serum bilirubin, mg/dL	82	1.40 (0.70–3.93)	0.70 (0.41–1.20)	<0.001
INR	79	1.20 (1.09–1.36)	1.10 (1.00–1.20)	0.036
Serum creatinine, mg/dL	73	1.01 (0.70–1.63)	0.70 (0.60–0.86)	<0.001
MELD score	64	9 (7–13)	8 (7–10)	0.081
ISGLS, n, %	79	4 (10)	35 (9)	0.780
Interstage interval, d	74	11 (8–14)	10 (8–14)	0.856
Stage 2 surgery				
Operation time, min	68	165 (140–215)	144 (110–200)	0.051
Concomitant resections, n, %	100	3 (6)	18 (4)	0.420

Categorical variables presented as count and percentage (%), continous variables presented as median and interquartile range (IQR).
*Defined as complications ≥3b; sFLR, standardized future liver remnant.
†Absolute change of sFLR volume before and after stage 1 surgery.
CVP indicates central venous pressure; BW, body weight; ICU, intensive care unit; INR, international normalized ratio.

only clinically useful parameters for decision making were finally included in the stepwise analysis purposely avoiding automatic variable selection. ''CRLM'' and ''chemotherapy before stage 1'' would be an example of 2 variables formally fulfilling the criteria but only CRLM were considered for multivariate analysis due to the clinical association of both variables. Since all CRLM underwent preoperative chemotherapy we wanted to avoid a false conclusion that chemotherapy itself is protective. Table 3 displays the respective logistic regression coefficients and odds ratios (ORs) with 95% confidence intervals (CIs) of pre-stage 1 and pre-stage 2 variables. Calibration of both regression models was assessed using the Hosmer-Lemeshow test.

Modeling Futile Outcome

Futility was modeled as function of identified risk factors in the multivariate regression analysis. Having 47 futile events from a total population of 528 patients, we aimed to study a maximum of 4 to 5 variables to prevent overfitting of the model as it is recommended in the 1-in-10 rule.[17]

A nonparametric receiver operating characteristic (ROC) analysis using the logit scores from the pre-stage 1 and pre-stage 2 final model was performed. The concordance (*c*) statistics was calculated as measure of model performance. For clinical practicability, cut-off variables were preferred over continuous variables. In this model cut-offs were, however, only used when the *c*-statistic was not inferior compared with the continuous alternative. The determination of the optimal cut-off value of continuous covariates (eg, age) was performed using ROC analysis and accuracy plots.[18] An ALPPS risk score formula for futile outcome was computed on the basis of the relative sizes of logistic regression coefficients (Table 3). In addition, we were interested if ISGLS liver failure at postoperative day (POD) 5 and complications 3b or higher are associated with the pre-stage 1 score before surgery.

All statistical analyses were performed using IBM SPSS Statistics version 22 for Macintosh (IBM Corporation, Armonk, NY).

RESULTS

Study Population

A total of 640 registered patients from 102 centers were screened to be included in the study. Of these 640, 146 patients (23%) were excluded due to various exclusion criteria such as centers reported less than 5 cases (n = 83), other ALPPS transection variants (n = 52), and failure to proceed with stage 2 (n = 11). Two centers outside the registry provided additional 34 patients with complete data sets. Finally, the study population was composed of 528 patients from 38 centers (Supplemental Fig. 1, http://links.lww.com/SLA/B60). The majority of ALPPS cases was performed for CRLM (69%) followed by 16% of cases for noncolorectal/nonbiliary and 15% for biliary tumors.

TABLE 3. Risk Modeling

	Risk Points	Regression Coefficient	Odds Ratio (95% CI)	*P*
Pre-stage 1 variables*				
Tumor type†				
CRLM (reference)	0	0.000	1.000	
Non-CRLM/nonbiliary	1	0.655	1.925 (0.808–4.585)	0.139
Biliary	2	1.326	3.767 (1.800–7.882)	<0.001
Age ≥67 yr	3	1.735	5.668 (2.843–11.30)	<0.001
Intercept pre-stage 1		−5.3		
Pre-stage 2 variables‡				
Pre-stage 1 score, per point	0.66	0.665	1.925 (1.527–2.426)	<0.001
Interstage complications ≥3b	1.2	1.209	3.350 (1.280–8.769)	0.014
Pre-stage 2 bilirubin§	1.5	1.496	4.439 (1.699–11.60)	0.002
Pre-stage 2 creatinine‖	1.7	1.696	5.454 (1.606–18.52)	0.007
Intercept pre-stage 2		−6.8		

*Pre-stage 1 risk score (range 0–5).

†Tumor type was classified into CRLM, biliary tumors, and non-CRLM/nonbiliary tumors with CRLM serving as reference. Scores of 0, 1, 2, 3, 4, and 5 were associated with futile risk of 2.7%, 4.9%, 8.6%, 15%, 24%, and 37%, respectively.

‡Pre-stage 2 risk score (range 0–12). Pre-stage 2 score = 0.66 × (pre-stage 1 score) + 1.2 × (1 = complications ≥3b; 0 = complications ≤3b) + 1.5 × $\log_{10}$ (10 × bilirubin pre-stage 2 in mg/dL) + 1.7 × $\log_{10}$ (10 × creatinine pre-stage 2 in mg/dL). Predicted pre-stage 2 futility risk is calculated according to following formula: Futility risk = odds/(1 + odds) with odds = exp (−6.9 + pre-stage 2 risk score).

§1.5 × $\log_{10}$ (10 × bilirubin pre-stage 2 in mg/dL).

‖1.7× $\log_{10}$ (10 × creatinine pre-stage 2 in mg/dL).

Peri- and Postoperative Outcome

The median operation time of stage 1 and 2 surgery was 305 (250–393 min) and 150 minutes (112–200 min). The major complication rate (grade ≥3b) and total complication rate after stage 1 surgery was 10% and 32%, respectively. The median volume increase of standardized FLR between stages was 70% (46%–106%). Approximately 11% of patients developed liver failure at POD 5 according to the ISGLS criteria. Stage 2 surgery was performed after a median interstage interval of 10 days (8–14 d) and the reported 90-day or in-hospital mortality rate was 9%.

Futile Versus Nonfutile Outcome

Among the 528 patients, 47 patients (9%) had a futile outcome after ALPPS surgery, with 18 patients (38%) and 45 patients (96%) dead at 1 and 3 months, 2 patients died in hospital after 3 months. Patients with futile outcome were in median 8 years older (69 vs 61 yr), had a higher proportion with CVD (21% vs 9%) and underwent less frequently resection for CRLM (41% vs 71%) than nonfutile cases (Table 1). Patients with a futile outcome experienced more frequently major postoperative complications after stage 1 surgery (33% vs 7%) and had a higher incidence of ISGLS-defined liver failure at POD 5 (23% vs 9%) (Table 2). The leading cause of death in the futile groups was septic shock in 18 patients (38%) followed by liver failure in 17 patients (36%), cardiogenic shock in 5 patients (11%), and other causes of death in 7 patients (15%).

Pre-stage 1 Prediction Model of Futile Outcome

The pre-stage 1 multivariate analysis identified age and biliary tumors as independent pre-stage 1 predictors for futility with the greatest risk from age 67 years or older (OR = 5.67; 95% CI 2.84–11.3), and biliary tumors (OR = 3.77; 95% CI 1.80–7.88) (Table 3). The calibration of the model was assessed by the Hosmer-Lemeshow test ($P = 0.235$). CVD marginally failed statistical significance ($P = 0.075$). ROC analysis and accuracy plot revealed age of 67 years as optimal age cut-off value for predicting futility (Fig. 1A, B). Age as continuous variable had no advantage over using the age cut-off of 67 years in predicting futility (*c*-statistic 0.771 vs 0.772). Using all 3 independent predictors, the futility prediction of the pre-stage 1 model had a *c*-statistic of 0.772 (95% CI 0.695–0.850) (Fig. 1C). A futility pre-stage 1 score was created on a 0 to 5 scale predicting futility risk of 2.7%, 4.9%, 8.6%, 15%, 24%, and 37% for patients with scores of 0, 1, 2, 3, 4, and 5, respectively (Table 3, Fig. 1D). There was a significant association of pre-stage 1 futility score and the probability of ISGLS liver failure at POD 5 ($P = 0.016$) and post-stage 1 complications 3b or higher ($P = 0.006$).

Pre-stage 2 Prediction Model of Futile Outcome

The pre-stage 2 multivariate analysis identified pre-stage 1 score (OR = 1.92; 95% CI 1.53–2.43), interstage complication grade 3b or higher (OR = 3.4; 95% CI 1.28–8.77), pre-stage 2 serum bilirubin (OR = 4.4; 95% CI 1.70–11.6), and creatinine (OR = 5.5; 95% CI 1.61–18.5), as independent pre-stage 2 predictors for futility with a Hosmer-Lemeshow calibration of $P = 0.909$ (Table 3). Using all 4 independent predictors, the futility prediction of the model had a *c*-statistic of 0.850 (95% CI 0.776–0.924) (Fig. 2A). The model predicted futility risk of 5%, 10%, 20%, and 50% for patients with scores of 3.9, 4.7, 5.5, and 6.9, respectively (Table 3, Fig. 2B).

DISCUSSION

The study identified age 67 years or older, biliary liver tumors, interstage occurrence of major complications, and elevated serum bilirubin or creatinine before stage 2 as independent predictors of futile outcome after ALPPS. When these predictors were integrated in a model predicting futile outcome after ALPPS, the discriminatory model performance was acceptable for the pre-stage 1 and good for the pre-stage 2 model. Both models provide statistical tools estimating the individual futile risk upfront stage 1 and before stage 2 surgery, which may assist the selection process whether and when to proceed with ALPPS surgery.

ALPPS is associated with many advantages in liver surgery but early mortality remains a serious problem for this newly practiced surgical technique. Even experienced hepatobiliary centers have reported high 90-day mortality rates ranging between 10% and 20%.[3–7] To avoid futile use of ALPPS, the main goal of the study

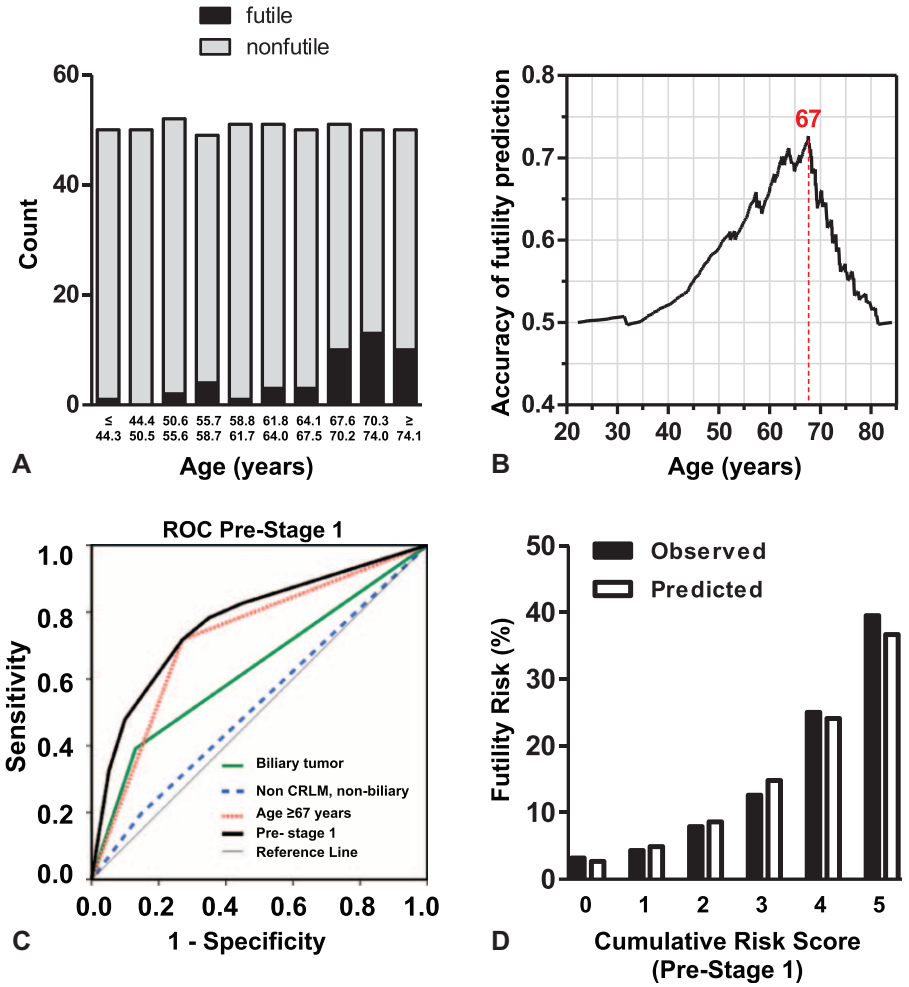


FIGURE 1. Pre-stage 1 prediction model of futile outcome. Percentage of patients experiencing futile outcome are plotted in age groups and show a marked increase beyond the age of 67 years (A). The accuracy plot over age as continuous variable identifies age of 67 years as optimal cut-off for predicting futility (B). C, ROC curves of single predictors of the pre-stage 1 model in comparison with pre-stage 1 prediction model (*c*-statistic 0.772; 95% CI 0.695–0.850). The pre-stage 1 model was composed of 2 predictors: age 67 years or older (*c*-statistic 0.723) and tumor entity (colorectal, biliary, noncolorectal/non-biliary). Modeling futility risk based on the pre-stage 1 cumulative risk score is illustrated in (D). The futility risk score is calculated according to the following formula: pre-stage 1 score = points age (0 = <67; 3 = ≥67 years) + points tumor (0 = colorectal; 1 = non-colorectal/non-biliary; 2 = biliary).

was to develop 2 prediction models, which would allow estimating the individual futile risks of patients before undergoing stage 1 and stage 2 surgery. In general, medical futility can be understood as treatment that extends life without providing quality of life or meaningful survival.[19] On the basis of this principle, we defined surgery-related futility when a patient did not reach a survival of 3 months or deceased during the postoperative course regardless of length of hospitalization. Cancer-related futility was not the objective of the present analysis and was not considered in the study.

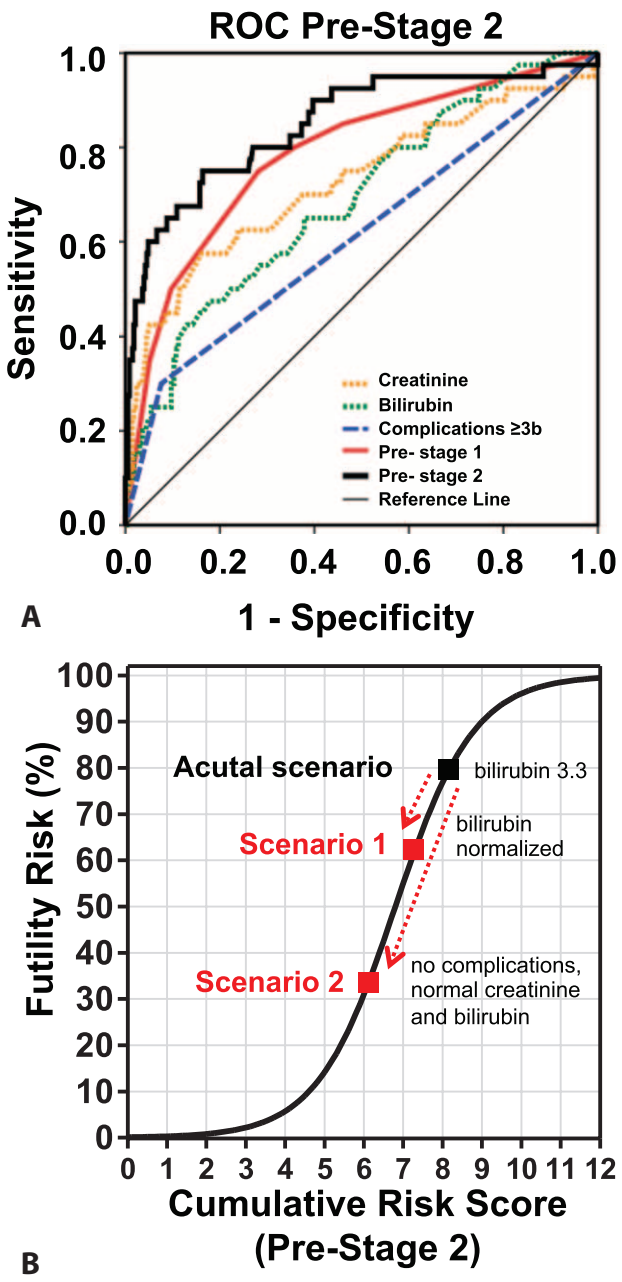
Age greater than 60 years has been repeatedly reported as risk factor for mortality after ALPPS.[8,9,20] In the present study, ROC and accuracy analysis identified age of 67 years as optimal cut-off for predicting futility. The reason for this discrepancy of age cut-offs might be related to the fact that previous studies[8–10] did not use statistical methods to look for the optimal cut-off point. The previous recommendation of age-related risk[20] needs to be revised because patients within the age segment of 60 to 67 years, which account for almost one-fourth of the study population (Fig. 1A), would be labeled with an incorrectly higher risk.

The study confirms previous findings that patients with CRLM have the lowest futile risk but identified biliary tumors carrying the greatest risk among all tumor types. Furthermore, the regression analysis identified the interstage occurrence of major complications and impaired renal and hepatic function before stage 2 surgery as additional independent predictors of futile outcome. The clinical effect of the interstage course on early mortality has been independently shown by previous studies.[7,8] Interestingly, measures of FLR volume and relative hypertrophy had no predictive ability in assessing futile risk in the present analysis.

The central part of the present study was to develop a risk model to estimate the likelihood of futile outcome before stage 1 and stage 2 surgery. When futility was modeled using the predictors age 67 years or older and tumor type, the pre-stage 1 risk model had an acceptable discriminatory ability with a *c*-statistic of 0.78. The use of age as continuous variable and the addition of CVD did not increase pre-stage 1 model performance. The pre-stage 2 model, which was based on the predictors pre-stage 1 score, inter-stage major complications and serum bilirubin and creatinine, had statistically superior performance over pre-stage 1 prediction with a *c*-statistic of 0.85. Models with *c*-statistics greater than 0.70 are considered clinically acceptable and greater than 0.80 are associated with good prediction. Values greater than 0.90 are rarely observed. Therefore, both risk models especially the pre-stage 2 model have a good performance in predicting futile risk in ALPPS. For reference, the commonly used risk assessment tool for estimating the 10-year risk of having a heart attack (Framingham Heart Study) had a *c*-statistic from 0.69 to 0.77, depending on the variables.[21]

Of note, the pre-stage 1 model presents a static risk score, whereas the pre-stage 2 score has a dynamic quality due to its modifiers serum bilirubin and creatinine. The applicability of both

risk models is demonstrated in Supplemental Table 1 (http://links.lww.com/SLA/B60), which displays clinical sample vignettes of 5 patients from the ALPPS registry. The example of the patient scenario in Figure 2B, however, demonstrates that even a normalization of the modifiers serum bilirubin and creatinine does not shift the elevated pre-stage 1 risk in a range to proceed safely with stage 2 surgery. Therefore, the previous recommendation that MELD scores less than 10 are the decision maker to proceed with stage 2 surgery needs to be reconsidered especially in patients with high pre-stage 1 futile risk.[8] The inclusion of composite variables such as MELD,[15] ISGLS,[16] and 50 to 50 criteria[22] into the model was avoided since these variables did not improve the predictive ability and are more susceptible for data incompleteness because more than 1 parameter are used.



The strength of the present study is related to the large sample size of 528 pooled patients from 38 centers. To reduce potential bias, we included only centers, which reported 5 or more cases to the registry. Another important advantage is that the primary endpoint of 90-day mortality presents a robust variable, which is reported by the vast majority of centers and does not require long-term follow-up. In contrast, the present study is also associated with certain inherent shortcomings, which are linked to a registry-based study design with higher degree of incomplete data sets. When data were initially retrieved from the registry, centers with higher degree of incomplete data were identified and contacted to provide missing data of important demographic and outcome variables. Another criticism of the registry study is the underreporting of negative events including early mortality. The observed 90-day mortality rate of 9% in the present series might not present the actual mortality figure of this procedure due to the potential bias of underreporting negative events to a registry-based database.[23,24] We therefore contacted centers in and outside the registry to report cases with futile outcome. By this approach, the initial events of futile outcome of 35 cases could be increased to 47 in the final study population. The greater sample size of futile events was of central importance for the study due to approaching a more solid statistical analysis.[17]

In conclusion, both risk models provide a mathematical tool to estimate the individual futile risk upfront ALPPS and before stage 2 surgery with a good discriminatory ability. These methods should not replace clinicians' best judgment, but might complementary assist proper patient selection and optimal determination whether and when to proceed safely with stage 2 surgery.

ACKNOWLEDGMENTS

The authors acknowledge the contribution of Livia Löffler and Sabine Kern from the Department of Surgery, University Hospital Zurich for data acquisition and data monitoring and Lisette Paratore-Hari, PhD from the Clinical Trial Center (CTC) Zurich for managing the electronic Case Report Form (eCRF) of the ALPPS registry.

FIGURE 2. Pre-stage 2 prediction model of futile outcome. A, ROC curves of single predictors of the pre-stage -2 model in comparison with pre-stage 2 prediction model (*c*-statistic 0.850 (95% CI 0.776–0.924)). The pre-stage 2 model was composed of 4 predictors: pre-stage 1 score (*c*-statistic 0.772), post-stage 1 complication ≥3b (*c*-statistic 0.612), and pre-stage 2 serum bilirubin (*c*-statistic 0.690) and creatinine (*c*-statistic 0.726). Modeling futility risk based on the pre-stage 2 cumulative risk score is illustrated in (B). The futility risk score is calculated according to the following formula: pre-stage 2 score = 0.66 × (pre-stage 1 score) + 1.7 × $\log_{10}$ (10 × creatinine mg/dL) + 1.5 × $\log_{10}$(10 × bilirubin mg/dL) + 1.2 × (1 = complication ≥3b; 0 = complication <3b). The predicted pre-stage 2 futility risk is calculated according to following formula: futility risk = odds/(1 + odds) with odds = exp(−6.9 + pre-stage 2 score). The squares on the futility risk curve represent patients #2 from Supplemental Table 1, http://links.lww.com/SLA/B60. The solid black square presents the actual risk, whereas the solid red squares present 2 scenarios of patient #2 with normalization of pre-stage 2 serum bilirubin (1 mg/dL) in scenario 1 and uneventful interstage course with normal renal (creatinine 1 mg/dL) and liver (bilirubin 1 mg/dL) function before stage 2 surgery in scenario 2.

REFERENCES

1. Schnitzbauer AA, Lang SA, Goessmann H, et al. Right portal vein ligation combined with in situ splitting induces rapid left lateral liver lobe hypertrophy enabling 2-staged extended right hepatic resection in small-for-size settings. *Ann Surg*. 2012;255:405–414.
2. De Santibanes E, Clavien PA. Playing Play-Doh to prevent postoperative liver failure: the ''ALPPS'' approach. *Ann Surg*. 2012;255:415–417.
3. Li J, Girotti P, Konigsrainer I, et al. ALPPS in right trisectionectomy: a safe procedure to avoid postoperative liver failure? *J Gastrointest Surg*. 2013;17:956–961.
4. Schadde E, Ardiles V, Slankamenac K, et al. ALPPS offers a better chance of complete resection in patients with primarily unresectable liver tumors compared with conventional-staged hepatectomies: results of a multicenter analysis. *World J Surg*. 2014;38:1510–1519.
5. Nadalin S, Capobianco I, Li J, et al. Indications and limits for associating liver partition and portal vein ligation for staged hepatectomy (ALPPS). Lessons Learned from 15 cases at a single centre. *Z Gastroenterol*. 2014;52:35–42.
6. Kremer M, Manzini G, Hristov B, et al. Impact of neoadjuvant chemotherapy on hypertrophy of the future liver remnant after associating liver partition and portal vein ligation for staged hepatectomy. *J Am Coll Surg*. 2015;221:717–728.
7. Truant S, Scatton O, Dokmak S, et al. Associating liver partition and portal vein ligation for staged hepatectomy (ALPPS): impact of the inter-stages course on morbi-mortality and implications for management. *Eur J Surg Oncol*. 2015;41:674–682.
8. Schadde E, Raptis DA, Schnitzbauer AA, et al. Prediction of mortality after ALPPS stage-1: an analysis of 320 patients from the international ALPPS registry. *Ann Surg*. 2015;262:780–786.
9. Schadde E, Ardiles V, Robles-Campos R, et al. Early survival and safety of ALPPS: first report of the International ALPPS Registry. *Ann Surg*. 2014;260:829–836.
10. D'Haese JG, Neumann J, Weniger M, et al. Should ALPPS be used for liver resection in intermediate-stage HCC? *Ann Surg Oncol*. 2015;23:1335–1343.
11. Linecker M, Kron P, Lang H, et al. Too many languages in the ALPPS: preventing another tower of Babel? *Ann Surg*. 2016;263:837–838.
12. Edmondson MJ, Sodergren MH, Pucher PH, et al. Variations and adaptations of associated liver partition and portal vein ligation for staged hepatectomy (ALPPS): many routes to the summit. *Surgery.* 2016;159:1058–1072.
13. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40:373–383.
14. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg*. 2004;240:205–213.
15. Malinchoc M, Kamath PS, Gordon FD, et al. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology.* 2000;31:864–871.
16. Rahbari NN, Garden OJ, Padbury R, et al. Posthepatectomy liver failure: a definition and grading by the International Study Group of Liver Surgery (ISGLS). *Surgery.* 2011;149:713–724.
17. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49:1373–1379.
18. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J*. 2005;47:458–472.
19. Helft PR, Siegler M, Lantos J. The rise and fall of the futility movement. *N Engl J Med*. 2000;343:293–296.
20. Oldhafer KJ, Stavrou GA, Van Gulik TM, et al. ALPPS-where do we stand, where do we go?: eight recommendations from the first international expert meeting. *Ann Surg*. 2016;263:839–841.
21. Wilson PW, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837–1847.
22. Balzan S, Belghiti J, Farges O, et al. The ''50-50 criteria'' on postoperative day 5: an accurate predictor of liver failure and death after hepatectomy. *Ann Surg*. 2005;242:824–828.
23. Dahl T, Rudjord K, Altreuther M, et al. Data quality of surgery for carotid artery stenosis. Are the national vascular registries reliable? *Eur J Vasc Endovasc Surg*. 2006;31:381–385.
24. Clavien PA, Puhan MA. Biased reporting in surgery. *Br J Surg*. 2014;101:591–592.

DISCUSSANTS

F. Pruvot (Lille, France):

Congratulations, the term ''futility score'' is inadequate and ''risk-score'' seems better. The risk of dying, in your article, range from 4.7% to 39% for the pre-stage 1 score and from 5% to 50% for pre-stage 2 score. Beyond which number do you think that ALPPS is ''futile,'' that is to say too risky or unnecessary?

We showed in Lille (EJSO 2015) that interstage course was as much important as was pre-stage 1 selection. But if the pre-stage 1 score may avoid a too risky procedure, the pre-stage 2 score may only help us to modulate the interstage course of patients. So, is the pre-stage 2 score only useful to postpone the stage 2? Do you think we can voluntarily abort the procedure? How behaved the 11 patients not having the stage 2 performed?

Interestingly, neither the baseline of the RLV nor its hypertrophy after stage 1 were predictive of mortality. Do you think that functional assessment using for instance hepatobiliary iminodiacetic acid (HIDA) scintigraphy may be particularly worth in firstly selecting good candidates for ALPPS (before stage 1) and secondly determining the best time point to perform the ALPPS stage 2?

According to my first reviewing, you listed some of the complications $\geq$3b after stage 1: biliary complications (15%), bleeding (13%), infection (6%). Did these local complications require re-intervention? Otherwise they could not be graded as 3b. Moreover, renal failure (27%) was the leading cause of severe interstage complications; thus, are the 2 variables ''major complications after stage 1'' and ''creatinine levels'' included in the pre-stage 2, strictly independent?

Modeling pre-stage 2 score, MELD and ISGLS were not contributive. Why didn't you try the 50/50 criteria?

Not distinguishing between the developmental period of ALPPS and the more recent period, in which technical improvements have been made, may be a major concern in the present study. Would reanalyzing the series with the more recent patients avoiding the complications of the learning curve, give another profile for the pre-stage 2 score?

Considering the number of events, the number of statistically significant variables, and the size of the cohort, shouldn't the authors try to use the Bootstrap method to enhance the robustness of their demonstration rather than wait another validation cohort that will take a long time to set up?

Response From H. Petrowsky (Zurich, Switzerland):

Thank you Dr Pruvot for your challenging and important questions. I will answer your questions sequentially.

Your first comment targets the question at which cut-off is ALPPS futile or too risky and should not be performed. The risk score is not an absolute recommendation of specific figures. The score should help patients and physicians to estimate the risk before stage 1 and stage 2 surgery. Both parties, patients and physicians, have to decide how much risk they want to take. Everybody in the auditorium would probably agree that a pre-stage 2 risk of 80% would permit proceeding to stage 2 surgery.

In your second question, you asked whether the pre-stage 2 score is only useful to postpone stage 2. As I pointed out in the presentation, the pre-stage 2 score can be used not only to postpone but also to deny stage 2 surgery, if the risk constellation is extraordinary high. I also want to emphasize again that previous recommendations, in which you just have to wait with stage 2 surgery until the liver function will be normalized, are not true for every case as I have shown in the clinical vignette that an extraordinary high risk

might be only marginally decreased by normalization of bilirubin and creatinine.

In terms of your third question, I agree with you that volume and hypertrophy were not predictive for mortality. Although the standardized future liver remnant was significantly lower in the futile group, it did not come out as independent predictor in the multivariate analysis. It might be that liver function is more important in predicting mortality than volume or hypertrophy. If this is true then liver function assessment by HIDA scan, as you pointed out, would be probably very helpful to reliably identify good candidates for ALPPS.

You also asked whether all major interstage complications required interventions? All complications graded as 3b or higher required intervention or intensive care unit management. We cannot exclude that renal failure as major complication has some minor degree of interaction with the risk factor creatinine but we think that this can be neglected because only one fourth of major complications are related to renal problems, and in addition, the risk factor creatinine reflects the pre-stage 2 renal function rather than the interstage course of creatinine.

Next, you asked why we did not try to use the 50–50 criteria? Indeed we did but the numbers of patients fulfilling the 50–50 criteria were very low that the inclusion of these criteria did not make sense from the statistical point of view.

Your question of the learning curve is a very important question. In general, the learning curve of a procedure or intervention can apply to technical improvements of the procedure itself or to selection and indication. We agree with you that technical improvements in ALPPS might have some effects on the learning curve, but more importantly we think that the learning curve in ALPPS is mainly associated with a modified selection of patients which are probably the main bulk of this analysis. But, to address the technical learning curve issue we included in our analysis only centers which entered at least 5 ALPPS cases in the registry.

In terms of your last question, we agree with you that bootstrapping could be another method to go. The main reason why we did not perform bootstrapping as internal validation method was that our goal was to validate the model externally in a future population of ALPPS cases entered in the registry.

C. Bruns (Cologne, Germany):

Last year, your group and others presented as far as I remember around 320 patients with also the idea to predict those patients who are not able to pass from stage 1 to stage 2. The final article, which was published in the November issue from last year, said that there definitely are some patients who should have no stage 2 and should be delayed or denied. We have to accept that ALPPS should be made safer. The mortality in that article, I think was approximately 8%, but that included those centers with less than 5 procedures. Now you have them excluded and we still end up with 9% mortality. That's my first question-please comment on this.

The second question is, in the last article there were already evaluated the MELD score, bilirubin, the age, and cardiac co-morbidity and the tumor entity. What is now the difference because that was in my opinion also a sort of prediction score? Could you comment on this?

Response From H. Petrowsky (Zurich, Switzerland):

Thank you for your important comments. I will start to answer the last question. Indeed, the study from last year showed some risk factors but it did not provide the cumulative risk of an individual patient scenario. The current project presents a further development of a pure regression analysis. What does it mean if the patient is older than 67 years and has a biliary tumor? Yes, the risk is elevated but we want to know the cumulative risk. In terms of data quality, we did a lot of efforts to improve the data completeness and quality. The article you mentioned from last year had 28 futile events, and in the present study 47. The previous article had a total of 320 patients; they present a total of 528 patients. Therefore, data quality and statistical analysis are more solid compared to the previous article.

Addressing you first question, yes it is true that the mortality is still 9%. The 2 centers that we contacted outside the registry were centers where we knew they had futile outcome. The problem with mortality is underreporting. Centers have a trend to report only good outcome and therefore we made every effort to include more futile events. Our goal was not primarily to increase the total number of the study population, but more importantly to increase the number of negative events with futile outcome to have a more solid statistical analysis.

R. Adam (Villejuif, France):

A first comment: You built a model of futility, and you exclude 11 patients who did not reach the second operation. By definition for me, not to reach the second operation is the best sign of futility. Thus, your model should take into account the 11 patients who did not reach the second stage.

My question concerns the clinical application of your risk assessment. What do you think could be the cut off for which you will decide to proceed or not to the second stage? Do you think that you will have a confrontation with the opinion of the oncologist for a chemotherapy that may help? Because of course, what do you do with a risk of 50%? Do you proceed or not?

Response From H. Petrowsky (Zurich, Switzerland):

Dr Adam, thank you for your critical comments and questions. Your first question was just addressed before. I agree that this could probably fit into the futility concept but just to have 11 patients represents a significant selection bias due to underreporting. I really recommend doing a study looking at patients who did not reach stage 2; however, this cannot be done by just the 11 patients reported to the registry.

Your second question targets the 50% risk. Yes, it is difficult but at the end, like in every other cancer surgery, you have to come up with a consent with your patient and your team including oncologists. What you also have to consider is that every prediction model has a predicting probability. For the pre-stage 2 model, the *c*-statistics was 0.85. This means that your prediction is correct in 85% but not in 15%. These numbers, compared to other prediction models, however, have a very good model performance. I think, at the end the prediction models might assist patient selection and make ALPPS safer.

T. van Gulik (Amsterdam, The Netherlands):

It will be extremely valuable to assess patients after stage 1 and be able to predict outcome after stage 2. What I missed is the kinetic growth rate of the remnant liver calculated after stage 1. This is an important indicator of the hypertrophy response. I did not see how you handled kinetic growth rate in your analysis.

Next, we have seen an evolution in the technical aspects of ALPPS. One of the things which certainly have decreased the operative risk is by performing a partial ALPPS at stage 1. How did you deal with this technical evolution in your data analysis?

Response From H. Petrowsky (Zurich, Switzerland):

Thank you, Thomas for your valuable comments on kinetic growth. We analyzed various volume variables but they did not come out as independent predictors. I also would like to make the point for

the community that kinetic growth can be only uniformly used if the volume scan is done for every patient at the same time. What we see in the registry is that there is a wide variation on the reported date of performing volume scans, for instance 40, 10, 8 days, and so on. If volume gain is divided by these time periods, you might get not the true growth rate knowing that there is no linear relationship over the entire period. Therefore, we have to be careful to use data on the kinetic growth rate. I think that this variable has to be completely controlled within a group and within a project. The registry will not provide this set of data.

In terms of technical evolution with partial ALPPS and ALPPS, you are right. We included partial transection variant to which we completely switched in Zurich in our analysis. At our center, we have good experiences with having a less invasive concept to do this procedure. We also looked at our current analysis but partial ALPPS or other variants did not come out as significant predictors.